



Deep Learning in Radiation Oncology

Wonjoong Cheon¹, Haksoo Kim¹, Jinsung Kim²

¹Proton Therapy Center, National Cancer Center, Goyang, ²Department of Radiation Oncology, Yonsei Cancer Center, Yonsei University College of Medicine, Seoul, Korea

Received 10 August 2020
Revised 27 August 2020
Accepted 3 September 2020

Corresponding author

Haksoo Kim
(haksoo.kim@ncc.re.kr)
Tel: 82-31-920-1757
Fax: 82-31-920-0149

Corresponding author

Jinsung Kim
(jinsung@yuhs.ac)
Tel: 82-2-2228-8118
Fax: 82-2-2227-7823

Deep learning (DL) is a subset of machine learning and artificial intelligence that has a deep neural network with a structure similar to the human neural system and has been trained using big data. DL narrows the gap between data acquisition and meaningful interpretation without explicit programming. It has so far outperformed most classification and regression methods and can automatically learn data representations for specific tasks. The application areas of DL in radiation oncology include classification, semantic segmentation, object detection, image translation and generation, and image captioning. This article tries to understand what is the potential role of DL and what can be more achieved by utilizing it in radiation oncology. With the advances in DL, various studies contributing to the development of radiation oncology were investigated comprehensively. In this article, the radiation treatment process was divided into six consecutive stages as follows: patient assessment, simulation, target and organs-at-risk segmentation, treatment planning, quality assurance, and beam delivery in terms of workflow. Studies using DL were classified and organized according to each radiation treatment process. State-of-the-art studies were identified, and the clinical utilities of those researches were examined. The DL model could provide faster and more accurate solutions to problems faced by oncologists. While the effect of a data-driven approach on improving the quality of care for cancer patients is evidently clear, implementing these methods will require cultural changes at both the professional and institutional levels. We believe this paper will serve as a guide for both clinicians and medical physicists on issues that need to be addressed in time.

Keywords: Artificial intelligence, Deep learning, Machine learning, Radiation oncology

Introduction

Deep learning (DL) is a subset of the larger family of machine learning technologies. Modern DL applies artificial neural networks (ANN) that use representation learning. The “deep” aspect in DL pertains to its application of multiple layers in a network, which resembles the human neural system. DL is not a novel technology, as it originated in brain science fields (e.g., neuroscience, neural engineering, and neurobiology). With the vast improvement and development in hardware performance, researchers wanted to

build computers that think like humans [1-3].

In 1950, Turing [4] was the first to formally ask the question “can machines think?” He also produced several important criteria for assessing machine intelligence. Walter Pitts and Warren McCulloch were the first to propose a Thresholded Logic Unit mimicking a neuron [5]. Soon after, the word of artificial intelligence was introduced to attendees by McCarthy at the Dartmouth Conference in 1956 [6]. In 1959, Rosenblatt demonstrated IBM’s Mark 1 perceptron, used for image recognition and classification [7]. The perceptron’s behavior was similar to the DL models of today.

In the case of Mark 1, photocells were adjusted by attached motors as part of a learning process to recognize US Mail postal codes.

However, the development of DL has stagnated for two periods: 1973–1980 and 1987–1993 [8,9]. However, it regained its momentum with the introduction and application of nonlinear activation functions [10], such as parallel processing. In 2012, modern DL was codified with AlexNet [11], which achieved significant milestones in machine learning perceptron performance with the graphics processing unit. By 2016, ResNet-200 [12], a DL model based on convolutional neural networks (CNN), finally surpassed the average human's score in image recognition and classification. Fig. 1 displays the advancements of DL computer in visual performance from 2011 to 2020.

DL has revolutionized several academic and industrial areas, including the medical field. The DL technique achieves superior recognition performance because it automatically extracts optimal features of images to produce learned classifications instead of relying on user-defined handcrafted features.

DL models can be classified into four structures that work effectively according to the problem type and data to be applied: multilayer perceptron (MLP), CNN, recurrent neural network

(RNN), and generative adversarial network (GAN) [13].

MLP and RNN are suitable for solving regression problems. Moreover, RNNs efficiently handle continuous data input, such as patient respiratory patterns and natural language processing tasks, due to their recursion capability. RNNs are augmented by long short-term memory (LSTM) [14], peephole connections [15], gated recurrent units [16], bidirectional LSTM (Bi-LSTM) [17], multiplicative LSTM [18], and LSTMs with attention [19].

CNN is widely used in analyzing visual imagery. It is comprised of several layers of convolution filters that are sometimes used in connection with MLP. The convolution filters are initialized randomly and optimized to achieve learning purposes. CNN is a shift- or space-invariant ANN; therefore, it is suitable for object detection and recognition tasks.

GANs are structurally used to generate new data or compare information across different domains, for example, from magnetic resonance imaging (MRI) to computed tomography (CT). GANs use a discriminator and a generator: the generator yields new data and the discriminator determines whether the newly created data are real or fake. Therefore, when the probability that the discriminator distinguishes newly generated data is 0.5, the training procedure is completed.

As, in recent years, DL in medical physics has evolved rapidly, medical physicists face the unavoidable task of translating this technology into the medical radiation oncology field. Radiation therapy is performed using high-energy radiations to deliver energy to the tumor [20]. Radiation therapy uses high-energy radiations to deliver energy to the tumor. To maximize tumor control probability (TCP) and minimize the normal tissue complication probability (NTCP), there are various radiation treatment processes as follows: (1) patient assessment, (2) simulation, (3) tumor and organs-at-risk (OARs) segmentation, (4) treatment planning, (5) quality assurance (QA), (6) beam delivery.

The current paper provides a succinct but comprehensive understanding of the great potentiality of DL and the corresponding roles of medical physicists. PubMed (<https://pubmed.gov/>) and the arXiv database (<https://arxiv.org/>) were utilized to search for published papers on DL for medical physics and radiation oncology from 2014 to 2020. Each study was categorized according to the subject.

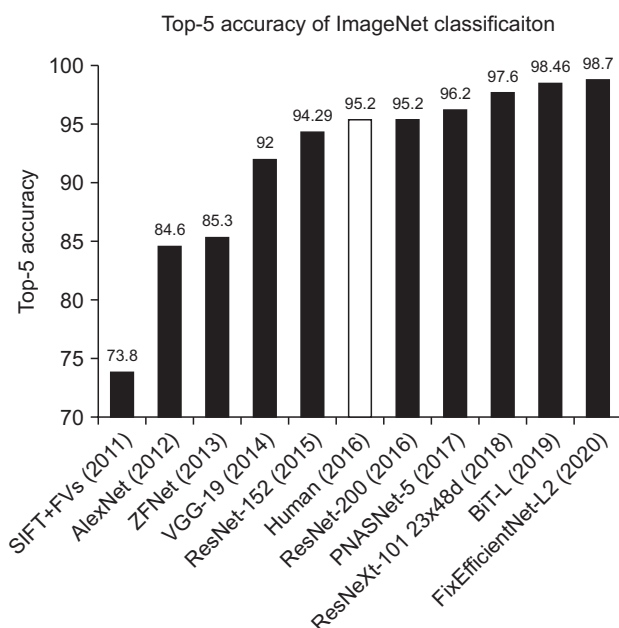


Fig. 1. Top accuracies for image classification models in ImageNet competitions over time.

Patient Assessment

1. Respiratory signal prediction

The position of target and OARs oscillate with patient's breathing pattern [21]. Thus, the internal target volume containing the tumor becomes larger and smaller, repetitively. Radiation therapy without taking into consideration the patient's respiratory pattern could lead to unnecessary radiation exposure, increasing NTCP [22].

To perform image-guided radiation therapy (IGRT) [23] or real-time tumor-tracking radiotherapy [24], according to the patient's breathing, understanding the movement patterns and trajectories of moving tumors and predicting their motion are essential. This is because radiation delivery systems generally have a latency of 50–150 ms. Moreover, a respiratory signal pattern prediction is necessary when conducting stereotactic radiosurgery (and stereotactic body radiotherapy and ultra-high dose rate (FLASH) radiotherapy technique that delivers 40 Gy or more per second [25].

Predicting the respiratory signal pattern is a regression problem; therefore, DL models based on MLPs or RNNs are quite suitable for this problem (Fig. 2).

In 2017, Sun et al. [26] conducted a comparison study using a random forest algorithm, an MLP, and adaptive boost-

ing with MLP (ADMLP) with normalized root-mean-square error (nRMSE) and Pearson's correlation coefficient as metrics. As a result, ADMLP had the lowest average nRMSE and the highest Pearson's correlation coefficients of 0.16 and 0.91, respectively.

Wang et al. [27] evaluated the accuracy of respiratory signal prediction using Bi-LSTM, demonstrating a better respiratory prediction performance than the autoregressive integrated moving average (ARIMA), which is commonly used for time series analysis and ADMLP. The nRMSE was 0.521, 0.228, and 0.081 for ARIMA, ADMLP, and Bi-LSTM, respectively. Bi-LSTM recorded the best performance for respiratory pattern prediction.

By reviewing the basic structure of LSTM, it can be understood why LSTM and the variant LSTM model outperform other algorithms and DL models. LSTM consists of three gates (i.e., input, forgot, and output) and a structure that transfers the status of cells containing LSTM to the next cell. This structure allows the LSTM model to achieve excellent performance when predicting future data from past data.

2. Radiotherapy outcome prediction

Recently, strategies for cancer treatment were developed based on multidisciplinary care, including physical surgery,

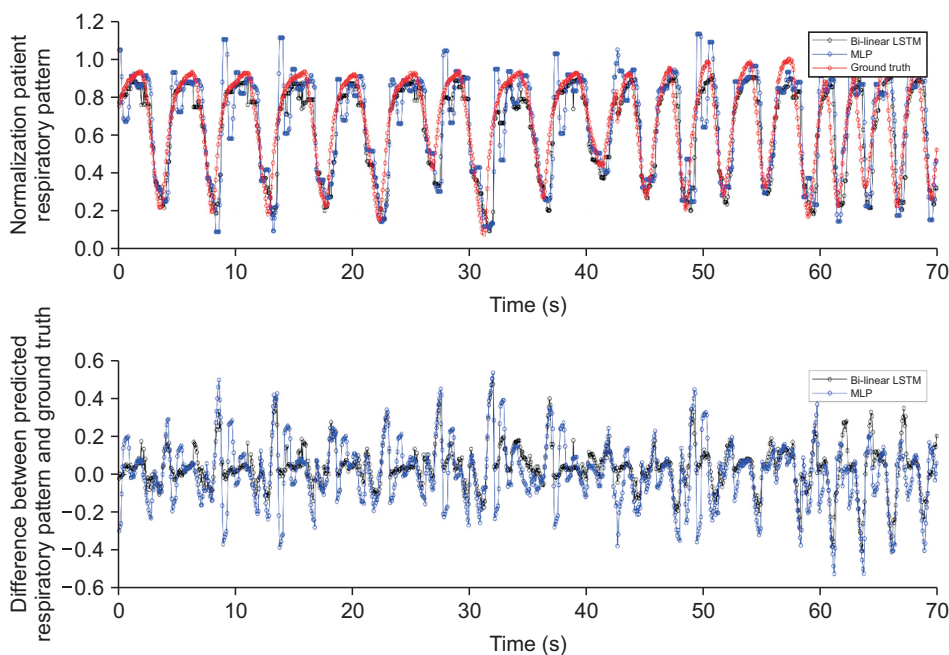


Fig. 2. Example of prediction of a patient's respiratory pattern using bilinear long short-term memory (LSTM; black), multilayer perceptron (MLP; blue), and ground truth (red).

chemotherapy, and radiation therapy.

About 30% of all patients with cancer in the Republic of Korea and 50% in the US have received radiation therapy. When starting radiation therapy, potential benefits should be assessed taking into account the TCP and the NTCP involved. The goal is to maximize TCP while minimizing NTCP [28]. For example, if the delivered absorbed dose to the tumor is extremely low, the treatment response decreases; or, if an unnecessarily high dose was delivered to the OARs, acute or late radiation toxicity (e.g., fibrosis or radiation therapy-induced oncogenesis) may occur. Thus, accurate risk assessment and prediction are essential, especially when alternatives such as physical surgery or chemotherapy are available.

The data given to perform radiation outcome prediction are divided into structured and unstructured [29]. The structured data (i.e., tabulated data) refer to data having intrinsic meanings, such as dosimetric, clinical, and biological variables; thus, a DL model based on the MLP or RNN family is recommended when building an outcome prediction model using structured data only. On the other hand, in the case of unstructured data, such as medical images or notes, a feature extractor is needed to extract meaningful information; therefore, CNN is generally recommended.

Arefan et al. [30] proposed a CNN-based two-class DL model with two schemes for predicting breast cancer risk. The first scheme was a pretrained CNN (GoogLeNet [31]) using the ImageNet dataset for deep feature extraction, whereas the second one was a CNN combined with a linear discriminant analysis (GoogLeNet-LDA) classifier. As a result, when the images of the whole breast were used as input data, the average area under the curve was 0.60 and 0.73 for GoogLeNet and GoogLeNet-LDA, respectively.

Li et al. [32] developed a CNN-based DL model to predict the survival risk in patients with rectal cancer. The prediction accuracy of the CNN model was compared with the random forest algorithm and Cox's proportional hazards model. The input data included CT, positron emission tomography (PET), and PET/CT combined images. Concordance-index (c-index) was used to assess the prediction performance obtained by different methods. As a result, the prediction accuracy of survival risk was the highest when the PET/CT combined images were used as input. The c-

index was 0.58, 0.60, and 0.64 for the random forest algorithm, Cox's proportional hazards model, and proposed CNN, respectively.

The CNN achieves higher performance than the other algorithms because of the advantages of DL. The DL model automatically extracts optimal features from the input to achieve the aim of the model. Although the analytical aspects of the features are challenging, they enable high performance.

Simulation Computed Tomography

High-quality simulated 3-dimensional (3D) CT images are essential when creating radiation treatment plans because the electron density and anatomical information of tumors and OARs are required to calculate and optimize dose distributions. Converting the Hounsfield unit (HU) to electron density is carried out to determine the accurate dose. Therefore, in radiology oncology, the simulated CT images are obtained using a CT simulator with a relatively larger bore size than that of a diagnostic CT, which requires a flatbed rather than the rounded one.

Studies on synthetically simulated CT image generation using DL can be divided into two types, according to the purpose: MRI-only radiotherapy and adaptive radiotherapy (ART). In the case of synthetic CT generation, CNNs or GANs are recommended, because they have shift-invariant and nonlinear characteristics.

MRI does not use ionizing radiation and has a relatively high soft-tissue contrast; therefore, relatively accurate target and OAR segmentations are possible. Currently, radiation oncologists use MRI/PET images to accurately segment a target on a simulated CT image.

If the contours of the target and OARs were drawn on the MRI and were transformed into a simulated CT image using image registration algorithms (e.g., deformable image registration) [33], an error could occur during the registration process. If MRI can be directly converted to simulation CT images without geometric distortions, MRI-only radiation therapy is possible.

Qi et al. [34] investigated a GAN-based DL model to generate synthetic CT images from MRI-based images for head and neck MRI-only radiotherapy. Different magnetic

resonance sequences and their combinations were tested to find optimal solutions. Consequently, the model with multiple magnetic resonance sequence images (T1, T2, T1 contrast, and T1DixonC-water) showed the best accuracies.

ART is a radiation therapy process, wherein the adopted treatment accounts for internal anatomical changes. With the current treatment processes and techniques [35], offline ART can be performed, which is time- and labor-intensive. To perform online ART, in which the patient is tracked by the patient positioning system, CT images considering the anatomical changes are required, which could be easily obtained as modern radiotherapy machines use cone beam CT (CBCT) to perform accurate positioning and IGRT. However, the CBCT is not suitable for dose calculation or adaptive planning, owing to the cupping and scattering artifacts and the inaccurate and unstable HUs [3,4]. Nevertheless, if CBCT can be converted into a simulated CT image using a DL model, the prerequisite to the online ART can be prepared.

Chen et al. [36] proposed a CNN-based DL model for generating simulated CT from on-treatment CBCT for patients with head and neck cancer. The mean absolute error (MAE) of HUs between CBCT and simulated CT was 44.38, and the HU difference between them was reduced to 18.89. Thus, the generation of synthetic CT from CBCT using CNN was verified.

As implied, CNNs can generate synthetic CT images with high accuracy. However, when leveraging CNNs or GANs, one must be careful when building a dataset. Efforts should be made to minimize the patient's physical changes when acquiring images using other imaging mechanisms to avoid errors related to the mismatches between images.

Target and Organs-at-Risk Segmentation

In the case of radiation therapy, the prescribed dose to the tumor is defined as the maximum and mean absorbed dose to the target volume or reference point. The dose limit for protecting OARs is the maximum and mean absorbed dose to an OAR volume. Therefore, defining the volume of the target and OARs is necessary to generate a treatment plan for radiation therapy.

The most time-consuming part of radiation treatment

planning is the target and OARs segmentation on the CT images. Thus, accurate and fast autosegmentation techniques are needed to reduce the patient's waiting time and to enable ART.

Segmentation consists of two tasks: recognition and delineation. Autosegmentation requires finding features (i.e., recognition) from images and judging the areas based on those features (i.e., delineation). Therefore, CNN has been widely used and recommended as an automatic feature extractor that can find optimal features from images, whereas MLP is mainly used as a predictor to judge a region class using extracted features. However, when MLP is utilized as the predictor, spatial information is lost and much more memory is required for the computation. Therefore, the trend is to use a fully convolutional network [37] consisting only of convolution layers instead of CNNs and MLPs [38].

In the field of medical image segmentation, diversity and accuracy of related research have grown rapidly since U-Net [39] was developed. U-Net is a CNN with an encoder structure that extracts features from images and a decoder structure that recovers the extracted features as a full-size segmentation map (Fig. 3). The concept of skip connections was also proposed [39], which provides local information to global information while upsampling.

Rachmadi et al. [40] automatically segmented white matter hyperintensities using a CNN model. They compared and evaluated each segmentation using a deep Boltzmann machine (DBM), support vector machine (SVM), random forest, and public toolbox comprising a lesion segmentation tool. Their proposed CNN model performance metric leveraged the dice similarity score (DCS), achieving the highest accuracy, followed by the DBM and random forest.

Zhu et al. [41] proposed a CNN model for fully automated whole-volume segmentation of head and neck patients, using MICCAI 2015 competition data. The segmented anatomies included brain stem, chiasm, mandible, optic nerve, parotid gland, and submandibular glands. AnatomyNet increased the DCS by 3.3% on average, providing the highest score in the previous competition.

Ahn et al. [42] conducted a comparative study for atlas- and DL-based autosegmentation of organ structures in liver cancer. The CNN model was FusionNet [43], using 70 cases with four OARs (i.e., heart, liver, kidney, and stomach). As

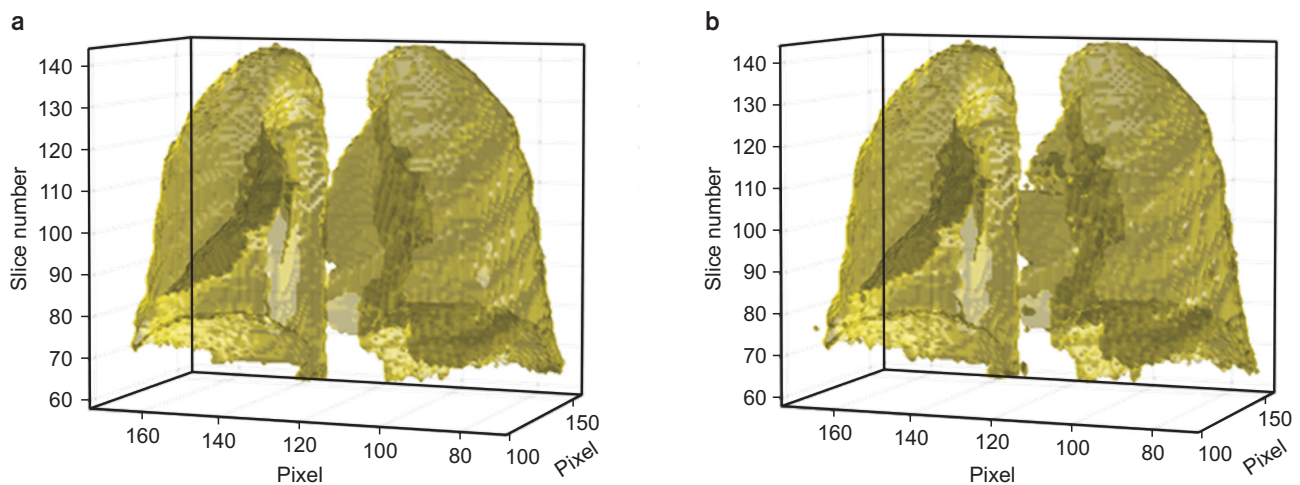


Fig. 3. Example of a 3-dimensional lung volume of (a) manual segmentation and (b) DL-based autosegmentation using U-Net.

a result, their DL-based model was superior to the atlas-based framework with a DCS of 3.6.

The most important activity in the autosegmentation task is defining the ground truth used to train the DL model. When building the dataset and training the DL models, the purity of the data is critical, known as “garbage-in, garbage-out.” In the target and OAR structure data, interobserver variability exists and must be recognized and handled [44-48].

Treatment Planning

1. Beam angle optimization

The beam angle configuration is a major planning decision which is constrained by the planner’s experience or template-based [49,50]. To automatically find an optimal beam angle while considering the dosimetric effect, generating candidates for the beam angle and optimizing a fluence map for all candidates to determine the optimal beam angle could be regarded as the problem. However, the difficult aspects of the beam angle optimization problem make it very challenging to simultaneously formulate it using a closed-form expression, which is computationally expensive because two-step optimization must be performed each time.

Recently, studies on beam angle optimization using a powerful DL algorithm have been published. Taasti et al. [51]

proposed a Bayesian optimization-based beam angle selection method in their in-house treatment planning system for pencil beam scanning. Bayesian optimization was used because nonconvex object functions can also be optimized.

Sadeghnejad Barkousaraie et al. [52] developed a CNN model that performed beam angle optimization. The CNN model trained using the results of the column generation method was used to carry out beam angle optimization to omit fluence optimization, which is computationally time-consuming.

Because volumetric arc therapy has become increasingly popular because of its high plan quality and efficient plan delivery [53,54], beam angle optimization may seem less appealing. However, with the advancements in proton and carbon therapies, beam angle optimization is still a relevant research area requiring further study.

2. Dose prediction

In the current radiation treatment planning procedures, the beam angle configuration is set by the planner, and the doses delivered to the target and OAR are optimized under the selected beam angle conditions. However, this process is very time-consuming and labor-intensive.

If a radiation oncology department has a variety of radiation therapy machines (e.g., medical linear accelerator, tomotherapy, or proton therapy), one must choose which machine to be used in the patient’s treatment. The best way



Fig. 4. Dose prediction for breast case: (a) optimized dose distribution by the treatment planning system, (b) predicted dose distribution by the deep learning model, and (c) dose difference between the optimized and predicted dose distributions.

is to create rival plans for each therapy type and compare the dose distributions. However, manually creating rival plans for all treatment devices is practically impossible. If the dose distribution reflecting the characteristics of each radiation therapy machine can be predicted using a DL model, it would help with planning and QA (Fig. 4).

Chen et al. [55] proposed a method for predicting optimal dose distributions, given the CT image and DICOM radiation therapy structure file using a CNN model (ResNet-101). They compared the accuracy of 2-dimensional (2D) dose distribution prediction based on input data. There are two input methods: one integrates the images and the radiation therapy structure; the other integrates the images, the radiation therapy structure, and the beam geometry. As a result, when beam geometry was included in the input, the predicted dose-volume histogram (DVH) was most similar to the correct DVH.

Barragán-Montero et al. [56] investigated the 3D dose distribution prediction method using a CNN model. They compared the predicted dose distributions using the anatomy-and-beam (AB) and the anatomy-only (AO) models. The two models predicted the dose distributions in the target volume with equivalent accuracy, resulting in a homogeneity index (mean \pm SD) of 0.11 ± 0.02 and 0.08 ± 0.02 for the AO and the AB models, respectively. In the case of the isodose volume in the medium-to-low dose region, the AO model was 10% less accurate than the AB model.

The biggest limitation of these studies is that they predicted only the dose distributions without a beam configu-

ration to operate the radiation therapy machine. Therefore, even if the dose distribution satisfies various criteria, it could still be useless. We believe that in the future DL-based autoplanning will be possible as long as studies are underway to generate beam configurations via the predicted dose distribution.

Other Topics

This section discusses several papers that are not included in the radiation treatment process but are related to other medical physics issues (e.g., QA, superresolution, material decomposition, and 2D dose distribution deconvolution).

1. Quality assurance

Regarding DL-based QA, Galib et al. [57] developed a model for automatically identifying and quantifying deformable registration errors using a CNN. The model had an architecture basement as the 3D U-Net and classified registrations into good or poor classes. The three channel inputs of the model were fixed image, moving image, and the absolute difference between them, while the outputs were class (good or poor) and registration error indices. The model was well-trained and showed reasonable performance with test data.

Nyflot et al. [58] proposed a patient-specific QA model employing a CNN model. In their paper, two experiments

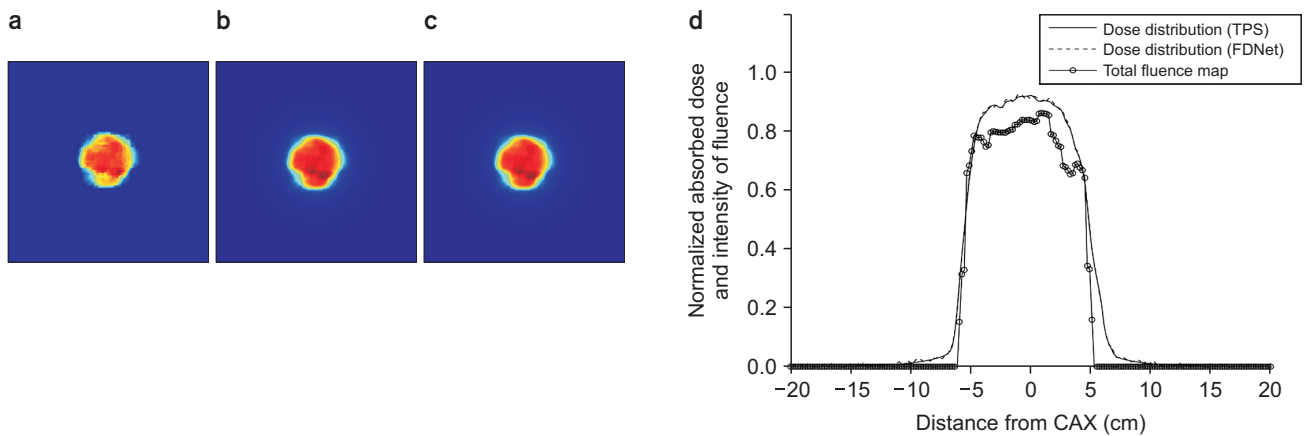


Fig. 5. Results of FDNet: (a) total fluence map, (b) dose distribution calculated by the treatment planning system, (c) predicted dose distribution using FDNet, and (d) profiles at the middle of the total fluence map, predicted and calculated dose distribution. TPS, treatment-planning system; FDNet, fluence-to-dose network; CAX, central axis.

were considered: a two-class experiment that classified images as error-free or containing a multileaf collimator (MLC) error and a three-class experiment classifying images as either error-free, containing a random MLC error, or containing a systematic MLC error. The CNN models were compared using four machine learning classifiers (i.e., SVMs, MLP, decision trees, and k -nearest neighbors). The highest accuracy was achieved using the DL approach with 77.3% and 64.3% maximum accuracies for two- and three-class experiments, respectively.

Interian et al. [59] developed a CNN model for predicting gamma passing rates of intensity-modulated radiation therapy (IMRT) plans from multiple treatment sites. The input of the CNN models included fluence maps reconstructed from the radiation therapy-plan file, while the output included gamma passing rates of the input plan. They compared the prediction accuracies of the proposed model and an ensemble of CNNs, where the MAEs were 0.70 ± 0.05 and 0.74 ± 0.06 for CNN and an ensemble of CNNs, respectively.

Cheon et al. [60] created a CNN model to predict the delivered dose distribution for patient-specific IMRT QA using a dynamic machine log file. The log file was reconstructed for a fluence stack, which was transformed to deliver the dose distribution of the proposed DL model (i.e., fluence-to-dose network [FDNet]; Fig. 5). The patient-specific IMRT QA was conducted using the proposed method, Gafchromic evidence-based therapy 3 (EBT3) film, and an ion chamber array detector. The average gamma passing rates were de-

termined using the 3%/3 mm gamma criterion. The results were 98.49%, 97.23%, and 98.03% for the proposed method, the EBT3 film, and the ion chamber array detector, respectively.

The advantage of performing QA using a DL model is that it can be performed without installing a QA device. However, because of the treatment machine conditions, including output, beam quality, symmetry, and flatness, and change over time, it is necessary to periodically reoptimize the DL-based QA model to maintain accuracy.

2. Superresolution

Kim et al. [61] proposed a CNN model for enhancing the image quality of MRIs incorporating another high-resolution MRI acquired using different MRI sequences. The input of models was low-resolution T2 sequence MRIs, whereas the output included high-resolution T2, T1, fluid-attenuated inversion recovery (FLAIR), and proton density sequence images for each model. The performance of the proposed model was compared using a compressed sensing (CS) algorithm for the evaluation metrics of nRMSE and a structural similarity index, revealing that the proposed model was superior to the CS algorithm.

Chun et al. [62] developed a DL GAN model to improve the image quality of a 3D low-resolution MRI for MRI-guided ART. The proposed superresolution generative (pSRG) model consisted of a denoising autoencoder, a downsam-

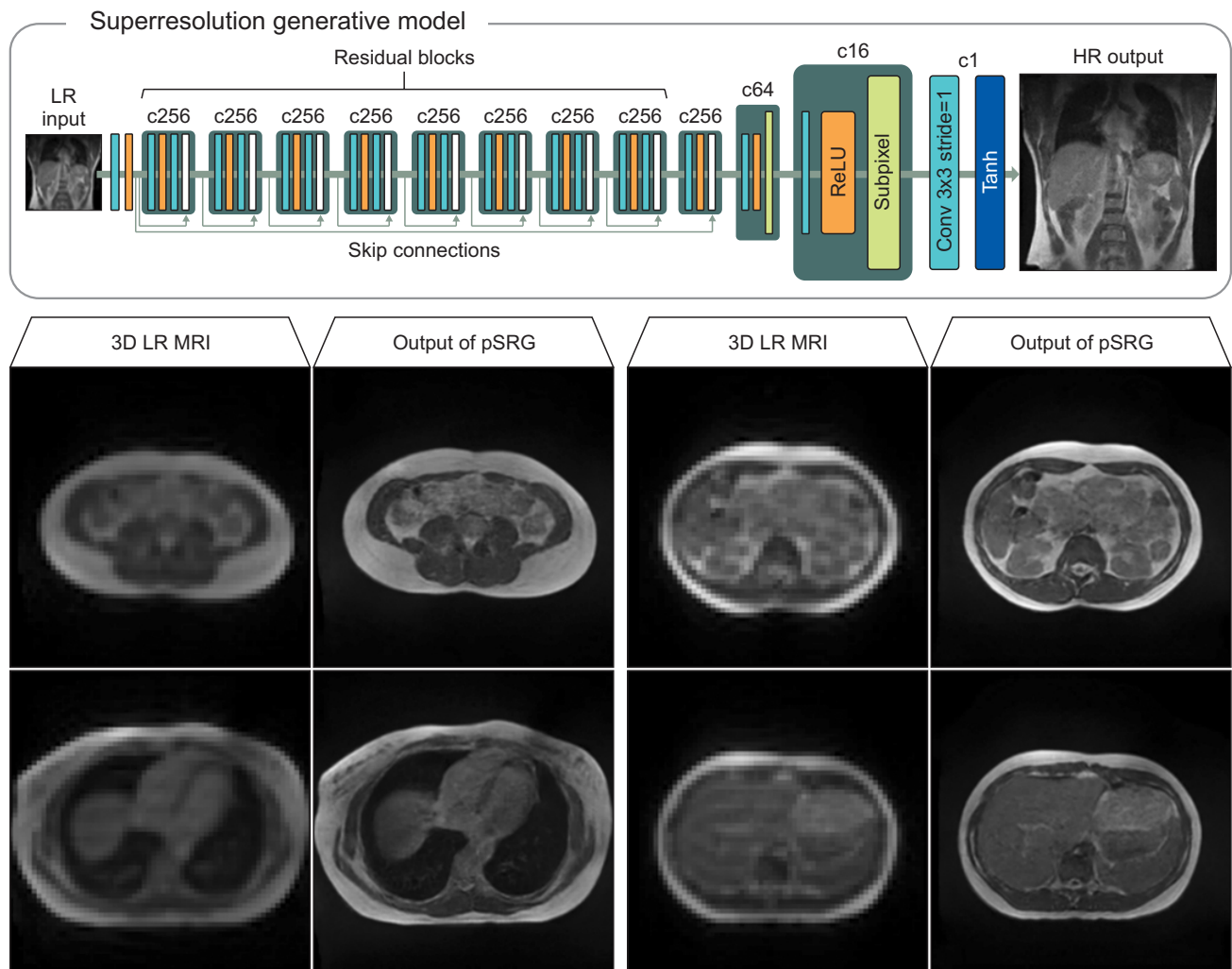


Fig. 6. Architecture and results of the proposed superresolution generative (pSRG) model for magnetic resonance imaging (MRI)-guided radiotherapy. LR, low resolution; HR, high resolution.

pling network, and a GAN. The high-resolution output of the pSRG model was compared to that of a conventional superresolution generative (cSRG) model using evaluation metrics of peak signal to noise ratio (SNR), Root Mean Square Error (RMSE), and a structural similarity index. pSRG showed better scores than those of cSRG in all evaluation metrics (Fig. 6).

Cheon et al. [63] proposed a CNN model to improve the image quality of a stereo portable gamma camera (SPGC) system designed to determine the position of the Bragg peak of a proton beam. The SPGC system detected proton-induced X-ray emissions generated from the interactions between the gold marker and a proton beam. To evaluate the performance of the proposed model, virtual experi-

ments were performed using the GEometry ANd Tracking 4 (GEANT4) package, where the in vivo proton range was measured using a standard SPGC system and another applying the proposed model. The averaged RMSEs of the five positions between the reference and calculation were smaller by 5.126 mm for the SPGC system applying the proposed model.

3 Material decomposition

In the field of radiation oncology, material decomposition can improve the accuracy of absorbed dose calculation by providing accurate material information. In the case of charge particles which have the characteristics of Bragg

peak, improving the calculation accuracy of the penetration depth of charged particle is possible.

Lu et al. [64] conducted a feasibility study for material decomposition using a CNN model. The performance was quantitatively assessed using a simulated extended cardiac-torso phantom and an anthropomorphic torso phantom. The accuracy of the proposed model was compared with the random forest method, where the proposed model exhibited better performance than the random forest by 4% and 16% in a noiseless and noisy environment, respectively.

4. Two-dimensional dose distribution deconvolution

When we performed dosimetry by using a dosimeter, the measured dose was influenced by the inherent characteristics of the measuring device: effective volume of the ion chamber, light sensitivity parameter of an image sensor of a scintillation detector, and so forth. If the deconvolution process was performed, the ground truth dose could be restored from the measurement dose.

Cheon et al. [65] developed a 2D dose distribution deconvolution network based on CNN for accurate 2D mirrorless scintillation dosimetry in the penumbra area. PenumbraNet, a model, was trained to correct the penumbra region of 2D dose distribution measured by an in-house scintillation detector. The performance of the PenumbraNet was then compared with an analytical deconvolution method based on Fourier theory. The gamma passing rate of the corrected 2D dose distribution was 11.04% higher than that of the analytical method when applying the 3%/3 mm gamma criterion.

Discussion and Conclusion

Radiotherapy plays an increasingly dominant role in the comprehensive multidisciplinary management of cancer [66]. As radiation therapy machines and treatment techniques become more advanced, the role of medical physicists, who ensure patients' safety, becomes more prominent.

With the advancement of DL, its powerful optimization capability has shown remarkable applicability in various fields. Its utility in radiation oncology and other medical

physics areas has been discussed and verified in several research papers [21-64]. These research fields range from radiation therapy processes to QA, medical image superresolution, material decomposition, and 2D dose distribution deconvolution.

This paper provides the trend of DL papers published thus far and serves as a tutorial and stepping stone for medical physicists.

Henceforth, medical physicists should be able to define the problems themselves, choose which DL models to use, collect data, perform appropriate preprocessing, and train and verify the DL models. Furthermore, commercial applications based on DL are becoming more widespread, and medical physicists will soon gain the ability to perform processes of acceptance and commissioning of DL-based applications.

Acknowledgements

Photographs courtesy of Sang Hee Ahn (National Cancer Center, Goyang), Jaehee Chun (Yonsei Cancer Center, Seoul), and Sang Woon Jeong (Samsung Medical Center, Seoul).

Conflicts of Interest

The author have nothing to disclose.

Availability of Data and Materials

All relevant data are within the paper and its Supporting Information files.

Author Contributions

Conceptualization: Haksoo Kim and Jinsung Kim. Data curation: Wonjoong Cheon. Formal analysis: Wonjoong Cheon, Haksoo Kim, and Jinsung Kim. Funding acquisition: Haksoo Kim and Jinsung Kim. Investigation: Wonjoong Cheon. Methodology: Wonjoong Cheon, Haksoo Kim, and Jinsung Kim. Project administration: Haksoo Kim and Jinsung Kim. Resources: Wonjoong Cheon, Haksoo Kim, and Jinsung Kim. Software: Wonjoong Cheon. Supervision:

Haksoo Kim and Jinsung Kim. Validation: Haksoo Kim and Jinsung Kim. Visualization: Wonjoong Cheon. Writing-original draft: Wonjoong Cheon. Writing-review & editing: Haksoo Kim and Jinsung Kim.

References

- Goodfellow I, Bengio Y, Courville A. Deep learning. Massachusetts: MIT Press; 2016.
- Bini SA. Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? *J Arthroplasty*. 2018;33:2358-2361.
- Garrido Á. Brain and artificial intelligence. *Brain*. 2017;8: 85-90.
- Turing AM. I.—Computing machinery and intelligence. *Mind*. 1950;59:433-460.
- McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biol*. 1943;5:15-133.
- McCorduck P. *Machines who think*. Natick: A K Peters, Ltd.; 2004.
- Rosenblatt F. *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. Washington D.C.: Spartan Books; 1961.
- Howe J. Artificial intelligence at Edinburgh University: a perspective. Edinburgh: the University of Edinburgh, 2007 [cited 2020 Jul 23]. Available from: <http://www.inf.ed.ac.uk/about/AIhistory.html>.
- Russell SJ, Norvig P. *Artificial intelligence: a modern approach*. 2nd ed. Upper Saddle River: Pearson Education, Inc.; 2003.
- Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. Paper presented at: ICML'10: Proceedings of the 27th International Conference on International Conference on Machine Learning; 2010 Jun 21-25; Haifa, Israel. p. 807-814.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. 2012:1097-1105.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Ithaca: arXiv.org, 2015 [cited 2020 Jul 23]. Available from: <https://arxiv.org/abs/1512.03385>.
- Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Ithaca: arXiv.org, 2014 [cited 2020 Jul 23]. Available from: <https://arxiv.org/abs/1406.2661>.
- Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9:1735-1780.
- Gers FA, Schmidhuber J. Recurrent nets that time and count. Paper presented at: the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium; 2000 Jul 27; Como, Italy. p. 189-194.
- Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. Ithaca: arXiv.org, 2014 [cited 2020 Jul 23]. Available from: <https://arxiv.org/abs/1406.1078>.
- Schuster M, Paliwal KK. Bidirectional recurrent neural networks. *IEEE Trans Signal Process*. 1997;45:2673-2681.
- Krause B, Lu L, Murray I, Renals S. Multiplicative LSTM for sequence modelling. Ithaca: arXiv.org, 2016 [cited 2020 Jul 23]. Available from: <https://arxiv.org/abs/1609.07959>.
- Wu Y, Schuster M, Chen Z, Le QV, Norouzi M, Macherey W, et al. Google's neural machine translation system: bridging the gap between human and machine translation. Ithaca: arXiv.org, 2016 [cited 2020 Jul 23]. Available from: <https://arxiv.org/abs/1609.08144>.
- Cui S, Tseng HH, Pakela J, Ten Haken RK, El Naqa I. Introduction to machine and deep learning for medical physicists. *Med Phys*. 2020;47:e127-e147.
- Hanley J, Debois MM, Mah D, Mageras GS, Raben A, Rosenzweig K, et al. Deep inspiration breath-hold technique for lung tumors: the potential value of target immobilization and reduced lung density in dose escalation. *Int J Radiat Oncol Biol Phys*. 1999;45:603-611.
- Kubo HD, Hill BC. Respiration gated radiotherapy treatment: a technical study. *Phys Med Biol*. 1996;41:83-91.
- Ramsey CR, Scaperroth D, Arwood D, Oliver AL. Clinical efficacy of respiratory gated conformal radiation therapy. *Med Dosim*. 1999;24:115-119.
- Shirato H, Shimizu S, Kunieda T, Kitamura K, van Herk M, Kagei K, et al. Physical aspects of a real-time tumor-tracking system for gated radiotherapy. *Int J Radiat Oncol Biol Phys*. 2000;48:1187-1195.
- de Kruijff RM. FLASH radiotherapy: ultra-high dose rates

- to spare healthy tissue. *Int J Radiat Biol.* 2020;96:419-423.
26. Sun WZ, Jiang MY, Ren L, Dang J, You T, Yin FF. Respiratory signal prediction based on adaptive boosting and multi-layer perceptron neural network. *Phys Med Biol.* 2017;62: 6822-6835.
 27. Wang R, Liang X, Zhu X, Xie Y. A feasibility of respiration prediction based on deep Bi-LSTM for real-time tumor tracking. *IEEE Access.* 2018;6:51262-51268.
 28. Chaikh A, Thariat J, Thureau S, Tessonnier T, Kammerer E, Fontbonne C, et al. [Construction of radiobiological models as TCP (tumor control probability) and NTCP (normal tissue complication probability): from dose to clinical effects prediction]. *Cancer Radiother.* 2020;24:247-257. French.
 29. Luo Y, Chen S, Valdes G. Machine learning for radiation outcome modeling and prediction. *Med Phys.* 2020;47:e178-e184.
 30. Arefan D, Mohamed AA, Berg WA, Zuley ML, Sumkin JH, Wu S. Deep learning modeling using normal mammograms for predicting breast cancer risk. *Med Phys.* 2020;47: 110-118.
 31. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. Ithaca: arXiv.org, 2014 [cited 2020 Jul 23]. Available from: <https://arxiv.org/abs/1409.4842>.
 32. Li H, Boimel P, Janopaul-Naylor J, Zhong H, Xiao Y, Ben-Josef E, et al. Deep convolutional neural networks for imaging data based survival analysis of rectal cancer. *Proc IEEE Int Symp Biomed Imaging.* 2019;2019:846-849.
 33. Tait LM, Hoffman D, Benedict S, Valicenti R, Mayadev JS. The use of MRI deformable image registration for CT-based brachytherapy in locally advanced cervical cancer. *Brachytherapy.* 2016;15:333-340.
 34. Qi M, Li Y, Wu A, Jia Q, Li B, Sun W, et al. Multi-sequence MR image-based synthetic CT generation using a generative adversarial network for head and neck MRI-only radiotherapy. *Med Phys.* 2020;47:1880-1894.
 35. Yang C, Liu F, Ahunbay E, Chang YW, Lawton C, Schultz C, et al. Combined online and offline adaptive radiation therapy: a dosimetric feasibility study. *Pract Radiat Oncol.* 2014;4:e75-e83.
 36. Chen L, Liang X, Shen C, Jiang S, Wang J. Synthetic CT generation from CBCT images via deep learning. *Med Phys.* 2020;47:1115-1125.
 37. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Ithaca: arXiv.org, 2014 [cited 2020 Jul 23]. Available from: <https://arxiv.org/abs/1411.4038>.
 38. Zhou Z, He Z, Jia Y. AFPNet: A 3D fully convolutional neural network with atrous-convolution feature pyramid for brain tumor segmentation via MRI images. *Neurocomputing.* 2020;402:235-244.
 39. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Ithaca: arXiv.org, 2015 [cited 2020 Jul 23]. Available from: <https://arxiv.org/abs/1505.04597>.
 40. Rachmadi MF, Valdés-Hernández MDC, Agan MLF, Di Perri C, Komura T; Alzheimer's Disease Neuroimaging Initiative. Segmentation of white matter hyperintensities using convolutional neural networks with global spatial information in routine clinical brain MRI with none or mild vascular pathology. *Comput Med Imaging Graph.* 2018;66: 28-43.
 41. Zhu W, Huang Y, Zeng L, Chen X, Liu Y, Qian Z, et al. AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Med Phys.* 2019;46:576-589.
 42. Ahn SH, Yeo AU, Kim KH, Kim C, Goh Y, Cho S, et al. Comparative clinical evaluation of atlas and deep-learning-based auto-segmentation of organ structures in liver cancer. *Radiat Oncol.* 2019;14:213.
 43. Quan TM, Hildebrand DGC, Jeong WK. FusionNet: a deep fully residual convolutional neural network for image segmentation in connectomics. Ithaca: arXiv.org, 2016 [cited 2020 Jul 23]. Available from: <https://arxiv.org/abs/1612.05360>.
 44. Vinod SK, Min M, Jameson MG, Holloway LC. A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. *J Med Imaging Radiat Oncol.* 2016;60:393-406.
 45. Rios Piedra EA, Taira RK, El-Saden S, Ellingson BM, Bui AAT, Hsu W. Assessing variability in brain tumor segmentation to improve volumetric accuracy and characterization of change. *IEEE EMBS Int Conf Biomed Health Inform.* 2016;2016:380-383.
 46. Apolle R, Appold S, Bijl HP, Blanchard P, Bussink J, Faivre-Finn C, et al. Inter-observer variability in target delineation increases during adaptive treatment of head-and-neck and lung cancer. *Acta Oncol.* 2019;58:1378-1385.

47. Chabane KTW. Interobserver variation of prostate delineation on CT and MR by radiation oncologists, radiologists and urologists at the Universitas annex oncology department [dissertation]. Bloemfontein: University of the Free State; 2018.
48. Franco P, Arcadipane F, Trino E, Gallio E, Martini S, Iorio GC, et al. Variability of clinical target volume delineation for rectal cancer patients planned for neoadjuvant radiotherapy with the aid of the platform Anatom-e. Clin Transl Radiat Oncol. 2018;11:33-39.
49. Cabrera GG, Ehrgott M, Mason AJ, Raith A. A matheuristic approach to solve the multiobjective beam angle optimization problem in intensity-modulated radiation therapy. 2018;25:243-268.
50. Breedveld S, Storchi PR, Voet PW, Heijmen BJ. iCycle: integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar IMRT plans. Med Phys. 2012;39:951-963.
51. Taasti VT, Hong L, Shim JSA, Deasy JO, Zarepisheh M. Automating proton treatment planning with beam angle selection using Bayesian optimization. Med Phys. 2020;47:3286-3296.
52. Sadeghnejad Barkousaraie A, Ogunmolu O, Jiang S, Nguyen D. A fast deep learning approach for beam orientation optimization for prostate cancer treated with intensity-modulated radiation therapy. Med Phys. 2020;47:880-897.
53. Quan EM, Li X, Li Y, Wang X, Kudchadker RJ, Johnson JL, et al. A comprehensive comparison of IMRT and VMAT plan quality for prostate cancer treatment. Int J Radiat Oncol Biol Phys. 2012;83:1169-1178.
54. Hoffmann M, Pacey J, Goodworth J, Laszczyk A, Ford R, Chick B, et al. Analysis of a volumetric-modulated arc therapy (VMAT) single phase prostate template as a class solution. Rep Pract Oncol Radiother. 2019;24:92-96.
55. Chen X, Men K, Li Y, Yi J, Dai J. A feasibility study on an automated method to generate patient-specific dose distributions for radiotherapy using deep learning. Med Phys. 2019;46:56-64.
56. Barragán-Montero AM, Nguyen D, Lu W, Lin MH, Norouzi-Kandalan R, Geets X, et al. Three-dimensional dose prediction for lung IMRT patients with deep neural networks: robust learning from heterogeneous beam configurations. Med Phys. 2019;46:3679-3691.
57. Galib SM, Lee HK, Guy CL, Riblett MJ, Hugo GD. A fast and scalable method for quality assurance of deformable image registration on lung CT scans using convolutional neural networks. Med Phys. 2020;47:99-109.
58. Nyflot MJ, Thammasorn P, Wootton LS, Ford EC, Chaovalitwongse WA. Deep learning for patient-specific quality assurance: identifying errors in radiotherapy delivery by radiomic analysis of gamma images with convolutional neural networks. Med Phys. 2019;46:456-464.
59. Interian Y, Rideout V, Kearney VP, Gennatas E, Morin O, Cheung J, et al. Deep nets vs expert designed features in medical physics: an IMRT QA case study. Med Phys. 2018;45:2672-2680.
60. Cheon W, Kim SJ, Hwang UJ, Min BJ, Han Y. Feasibility study of the fluence-to-dose network (FDNet) for patient-specific IMRT quality assurance. J Korean Phys Soc. 2019;75:724-734.
61. Kim KH, Do WJ, Park SH. Improving resolution of MR images with an adversarial network incorporating images with different contrast. Med Phys. 2018;45:3120-3131.
62. Chun J, Zhang H, Gach HM, Olberg S, Mazur T, Green O, et al. MRI super-resolution reconstruction for MRI-guided adaptive radiotherapy using cascaded deep learning: in the presence of limited training data and unknown translation model. Med Phys. 2019;46:4148-4164.
63. Cheon W, Lee J, Min BJ, Han Y. Super-resolution model for high-precision *in vivo* proton range verification using a stereo gamma camera: a feasibility study. J Korean Phys Soc. 2019;75:617-627.
64. Lu Y, Kowarschik M, Huang X, Xia Y, Choi JH, Chen S, et al. A learning-based material decomposition pipeline for multi-energy x-ray imaging. Med Phys. 2019;46:689-703.
65. Cheon W, Kim SJ, Kim K, Lee M, Lee J, Jo K, et al. Feasibility of two-dimensional dose distribution deconvolution using convolution neural networks. Med Phys. 2019;46:5833-5847.
66. Grau C, Hoyer M. High-precision radiotherapy. Eur Oncol Rev. 2005;40-44.